DOCUMENT RESUME

ED 346 139 TM 018 459

AUTHOR Ryan, Katherine E.
TITLE The Impact of Different Focal Group Specifications on DIF.
PUB DATE Apr 92
NOTE 21p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 21-23, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Advanced Placement; *Black Students; *College Freshmen; Comparative Testing; Economically Disadvantaged; Educationally Disadvantaged; Ethnicity; Group Membership; Higher Education; *Hispanic Americans; *Item Bias; *Mathematics Tests; Racial Bias; Racial Differences; Test Items; *White Students
IDENTIFIERS Advanced Placement Examinations (CEEB); African Americans; *Focus Groups Approach; Mantel Haenszel Procedure

ABSTRACT

Whether specification of focal group membership based on ethnicity as well as educational opportunity affects differential item functioning (DIF) of test items was studied. Subjects were 4,342 freshmen admitted to a large midwestern university who completed the 40-item Advanced Mathematics Placement Examination in the spring of 1991. There were 3,896 White, 338 African American, and 108 Hispanic American examinees. All of the Hispanic Americans and 166 of the African Americans were from high schools judged to be educationally and economically disadvantaged. The Mantel-Haenszel (MH) procedure was used to examine DIF. Although there were few verbal items, the mean MH DIF delta difference did change substantially in relation to focal group specification. There were also changes in item classification depending on the specification of focal group membership. The question of whether sample sizes for focal group measurement were adequate in this investigation is discussed. There are five tables of study data and one illustrative plot. A 23-item list of references is included. (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
KATHERINE E. RYAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED346139

# THE IMPACT OF DIFFERENT FOCAL GROUP SPECIFICATIONS ON DIF

Katherine E. Ryan
Division of Measurement and Evaluation
University of Illinois
307 Engineering Hall
1308 W. Green Street
Urbana, IL 61801

Paper presented at the annual meeting on the National Council on Measurement in Education, San Francisco, April 1992

TM018459



BEST COPY AVAILABLE

While there has been major progress in the development of empirical methods for detecting differential item functioning (DIF) (Dorans, 1989; Holland, 1985; Linn, Levine, Hastings, Wardrop, 1981; Swarminathan & Rogers, 1990), it has proven much more difficult to identify factors which contribute to DIF (Linn & Harnisch, 1981; Schmitt, Holland, & Dorans, 1989). One issue that may contribute to the difficulty is the notion of comparability. According to Holland and Thayer (1988), theoretically, comparable group members are to be matched on several characteristics relevant to performance on the item. These include performance on measures of the ability of interest, schooling, curricular background, additional pertinent group memberships, and other relevant characteristics. In practice, however, the matching of examinees is usually based on test scores on the ability of interest.

Closely related to the notion of comparability is the specification of the focal group membership. Focal group members are usually treated as a homogeneous group, specified on the basis of ethnicity or gender alone (Schmitt et al., 1989). However, the focal group members are likely to be more diverse (Schmitt et al., 1989). There are several investigations that have addressed comparability and the specification of the focal group in relationship to DIF. In studies examining DIF for Asian-Americans and Hispanics on the Scholastic Aptitude Test (SAT), when the focal group members included only test takers who indicated English was their best language, few items exhibiting large DIF were found (Bleistein & Wright, 1987; Schmitt, 1988; Schmitt & Dorans, 1990). In contrast, Zwick and Ercikan (1989) found that controlling on material studied and total test score did not decrease the number of DIF items on the U.S. History NAEP items on the NAEP in comparison to the number of DIF items controlling on total test score alone.

Linn and Harnisch (1981) suggested DIF may result from differences in educational experiences. Bonds (1989) and Ramsey (1989) have suggested that the focal and reference group membership be designated on the basis of educational advantage and disadvantage. The purpose of this investigation is to examine whether specification of focal group membership based on ethnicity as well as educational opportunity affects DIF results.

## Methods

### Sample

Subjects of this investigation were 4,342 freshmen admitted to a large midwestern university who completed the 40 item Advanced Mathematics Placement Examination in Spring, 1991. There were 3,896 white test takers; 338 examinees were African-Americans, and 108 examinees were Hispanics. Of the African-American and Hispanic freshmen, 166 of the African-Americans and all of the Hispanics in the sample were from high schools judged to be educationally and economically disadvantaged and participated in the university's Educational Opportunity Program (EOP). The criteria used by the university admissions in judging high schools are based on a) income levels of the population surrounding the high school; b) the percentage of minority students enrolled; and c) the rate of student turnover at the school.

The Advanced Mathematics Placement Examination is a four response option multiple choice test administered to students who have at least 3.5 years of high school mathematics or have completed a course in trigonometry. The mean test score from this administration is 21.94 and the reliability is =.94 (coefficient alpha) which is typical for this examination. Generally, there is a substantial difference in the item difficulties for white and African-American test takers (.56 for whites; .39 for African-Americans).

The exam consists of intermediate and advanced algebra items (45%), plane geometry items (20%), and trigonometry items (35%). Previous research suggests verbally-loaded mathematics items are differentially easier for white test takers (Camilli & Smith, 1990; Ryan, 1991; Shepard, Camilli, & Williams, 1985). Although, the majority (24 items) of items were mathematical sentences or diagrams, a scheme from previous research was used to categorize items (See Bleistein and Wright, 1987). Item types were classified in one of three categories: pure mathematics, verbally loaded, and neutral. In accordance with the Bleistein and Wright's scheme, mathematics items are primarily formula-based items involving minimal language or reading comprehension. Verbally-loaded items, in this context, are story problems presenting a scenario or alternatively, are textually-based items, but require mathematical reasoning skills to solve the problem. The third category, neutral items, are items that do not fit in either category. Of the forty test items, three were classified as verbal while thirteen were considered to be neutral.

Dimensionality

To examine test dimensionality, Reckase's (1979) approach was used. Tetrachoric inter-item correlations were obtained. A scree plot of the eigen values was obtained and inspected. The first eigen value was over six times as large as the second eigen value for the 40 items suggesting the dominance of the first factor (Hambleton, Swaminathan, & Rogers, 1991). As a consequence, no subtest scores were used as a matching criterion in the analyses.

Mantel-Haenszel Analyses

The Mantel-Haenszel procedure (Mantel & Haenszel, 1959) consisting of the Mantel-Haenszel common odds ratio and the Mantel-Haenszel chi-square statistic, and the MH delta difference (MH D-DIF) (Holland, 1985) were calculated to examine

DIF. The common odds ratio estimate is interpreted as the average factor by which the probability that reference group members are correct on an item is larger than the probability for comparable members of the focal group (Holland & Thayer, 1986). The MH chi-square has an approximate chi-square distribution with one degree of freedom under the null hypothesis of no DIF. Because the alternate hypothesis is nondirectional, the MH chi-square statistic can identify DIF that favors either subgroup (Holland & Thayer, 1988).

The MH D-DIF is the difference in item difficulty in the ETS delta metric for comparable reference and focal group members (Scheuneman & Gerritz, 1990). Items with a MH D-DIF equal to zero are of equal difficulty for matched groups of test-takers. A negative value for MH D-DIF indicates that the item is differentially easier for the reference group than for the focal group (Holland & Thayer, 1988). Positive MH D-DIF values indicate the item is easier for focal group members in comparison to comparable reference group members. In addition, the standard errors of the MH D-DIF (Phillips & Holland, 1987) were also obtained. Calculations of these indices are described in detail elsewhere (See Holland, 1985; Holland & Thayer,1986; Mantel & Haenszel, 1959).

Using Holland's (1985) terminology, in this investigation, examinees were classified as either focal group members (subgroup of interest) or reference group members (standard to compare performances of the focal group). Test takers were then matched on total test score and score groups were not collapsed. Thus, for the Advanced Mathematics Placement Examination (40 items), there were potentially 41 score groups. The item under investigation was included in the matching criterion (Holland & Thayer, 1986). Item classification was based on the criteria established by ETS to evaluate item functioning (Zwick & Ericikan, 1989).[1] "A" items are considered to be free from DIF. "B" items are also

candidate items; however, items with the lowest MH D-DIF values are to be selected. Items which are classified as "C" items are picked only to meet test specifications.

Design of MH Analyses

Table 1 provides a description of the MH analyses that were conducted. All analyses tested the 40 items from the Advanced Mathematics Placement Examination. Analyses 1 and 2 were carried out to provide a baseline. A spaced sample of 345 white examinees were selected as the focal group; the remaining white test takers were the reference group. These two groups constitute the WW sample for Analysis 1. To examine the effects of differences in score distributions, a baseline analysis, using 168 white test takers with a score distribution that is similar to the EOP African-American sample (N=166), was selected as the focal group. (This white focal group was selected by score group without replacement to match the African-American EOP sample). The rest of the white test takers were designated as the reference group. These two groups will form the WMW sample for Analysis 2.

---

Insert Table 1 here

---

The design of the analysis in Analysis 3 parallels that in Analysis 2, except that the entire African-American sample was designated as the focal group. The white test takers were specified as the reference group. All white and African-American examinees constitute the AAW sample. To investigate whether classifying test takers on the basis of ethnicity and educational opportunity affects DIF results, the EOP African-American students (who come from high schools that are designated as educationally and economically disadvantaged) were the focal group members for Analysis 4. The reference group was the same as in Analysis 3 (all white examinees); these two groups compose the EOPAW sample. To

study the impact of homogenous focal group membership specification (Analysis 5), the NAW sample consisted of non-EOP African-American students (students who are not designated as educationally and economically disadvantaged) as the focal group. All white test takers were designated as the reference group. Further, to examine the effects of classifying students based on educational opportunity, EOP students (Hispanic and African-American only) who come from high schools that are judged to be economically and educationally disadvantaged were designated as the focal group while the white test takers were specified as the reference group. These two groups comprise the EOPW sample for Analysis 6.

## Results

### Descriptive Statistics for the Advanced Math Exam by Group

Table 2 presents the descriptive test statistics. The mean test score for the entire sample of African-American test takers is approximately 1 standard deviation lower than white test takers. The mean test scores and standard deviations for the EOP African-American group and the entire EOP group are nearly identical. In contrast, the mean test score for the non-EOP African-American sample is approximately .25 standard deviations lower than the white sample and .75 standard deviations higher than the EOP African-American and entire EOP samples.

---

Insert Table 2 here

---

### Descriptive Statistics for the MH D-DIF Values

The means and standard deviations for the MH D-DIF for all the samples are presented in Table 3. The means for MH D-DIF values for the white comparisons (WW and WMW) are slightly lower than the means for the ethnic and EOP comparisons. The standard deviation for the matched white comparison (WMW) is larger than the

unmatched comparison (WW) and is similar in size to the white-African American (AAW) and the white-non EOP African-American comparisons. The standard deviation for the white-EOP African-American comparison is largest. Previous research suggests the size of the standard deviations for the EOPAW sample and EOPW sample is typical when ethnicity alone is used as a matching variable (Ryan, 1991; Wright, 1987). In contrast, the standard deviation for the NAW sample is smaller.

---

Insert Table 3 here

---

Item Categorization

The means and standard deviations for the MH D-DIF values by item category. are presented in Table 4. For the pure and neutral items, the means for all the samples are close to zero like the overall means for the MH D-DIF values. In contrast, the means for the verbal items tended to increase for the NAW sample and are substantially higher for the EOPAW and EOPW samples. However, these means are based on the MH D-DIF values for only three verbal items. Like the overall standard deviations for the MH D-DIF values, the standard deviations are largest for the EOPAW and EOPW samples. In general, the neutral items are slightly easier for comparable focal group members for these two samples.

---

Insert Table 4 here

---

Table 5 presents the results of the MH D-DIF analyses. As shown in Table 5, most of the items were found to be free from DIF for most of the samples. No items reached the criteria for a "C" classification. The samples controlling for educational opportunity (EOPAW and EOPW) had the largest number of "B" items (7 in each). However, not all flagged items were differentially easier for the reference group. Looking across all samples, approximately equal numbers of items

were differentially easier for the reference group and the focal group respectively.

For the first baseline analysis, all of the MH D-DIF values for the unmatched sample (WW) were classified as "A" items. For the matched white sample (WMW), two trigonometry items categorized as pure were differentially easier for the reference group (the large white sample).

For the AAW, NAW and EOPAW, and EOPW samples, two items were found to be easier for comparable focal group members. These items were categorized, respectively, as a pure algebra item (32) and a neutral trigonometry item (39). Item 18 was differentially easier for the white test takers in the AAW, EOPAW and EOPW samples, but did not meet criteria in the NAW sample.

Of the seven items were classified as "B" items for the EOPAW sample, three items were differentially more difficult for the focal group (EOP African-American test takers). Item 4 was categorized as a pure algebra item. Item 18 is a verbally-loaded algebra item while item 39 is a pure geometry item. Two of these items (18 and 39) were also flagged as easier for comparable reference group members in the EOPW sample. However, the EOPAW and EOPW samples overlap.

---

Insert Table 5 here

---

When comparing the MH D-DIF values for the NAW and EOPAW samples, there are considerable shifts for some of the items. Figure 1 presents a plot of the MH D-DIF values for the NAW and EOPAW samples. For example, item 18 has a MH D-DIF equal to -.654 in the NAW sample and a -2.028 in the EOPAW sample.

---

Insert Figure 1 here

---

## Discussion

This investigation examined whether designating focal group membership on the basis of educational opportunity and ethnicity impacts DIF. Although there were a small number of verbal items, the mean MH D-DIF value for the verbal items does change substantially in relationship to focal group specification. In addition, there are also changes in item classification depending on the specification of focal group membership. When considering only the items that were easier for comparable reference group members, the one item flagged in the traditional sample (AAW) was not in the NAW sample. In contrast, when focal group members were specified on the basis of educational opportunity and ethnicity, the number of "B" items increased. Two items which were classified as "A" items in the AAW and NAW samples met the criteria for "B" classification in the EOPAW sample, and were considered differentially easier for the reference group .

There were also a larger number of items found to function differentially, in general, in the EOPAW sample. Previous research has suggested the MH D-DIF values from the small samples are affected by sample characteristics (Ryan, 1991; Wright, 1987). The size of this sample is modest (N=166). However the WMW and NAW samples are similar in size.

The mean difference in performance between the focal and reference group members is substantially larger in the EOPAW sample than the NAW sample. Zwick (1989) has suggested that when there are differences in ability distributions, the MH procedure may detect DIF when none exists. The mean difference in performance was similar for the focal and reference group for the WMW and EOPAW samples. There were only two items flagged in the WMW sample. However, while the score distributions were comparable for these two focal groups, there may well be considerable differences in other characteristics such as educational

background and geographical region.

It is not clear that sample sizes for the focal group membership were adequate in this investigation (McPeek and Wild 1986; Wright, 1987). For a small testing programs, this is always an issue. Cross validation with a different sample of test takers from previous administrations is of interest. In addition, the classification scheme used in this investigation is limited. Neutral is a broad category. Research investigating properties of mathematical items that affect item difficulty are needed much like those that have been conducted in the area of verbal ability would be useful (Scheuenman & Gerritz, 1990).

References

Bleistein, C. A., & Wright, D. (1987). Assessment of unexpected differential item difficulty for Asian-American candidates on the Scholastic Aptitude Test. In A. P. Schmitt & N. J. Dorans (eds.), Differential item functioning on the Scholastic Aptitude Test (RM-87-1). Princeton, NJ:Educational Testing Service.

Bond, L. (1989, October). Discussant comments to Item and test characteristics associated with DIF. Paper presented at the Differential Item Functioning: Theory and Practice Conference at Educational Testing Services, Princeton, NJ.

Camilli, G., & Smith, J.K. (1990). Comparison of the Mantel-Haenszel test with a randomized and a jackknife test for detecting biased items. Journal of Educational Statistics, 15, 53-67.

Dorans, N. J. (1989). Two new approaches to assessing unexpected differential item functioning: Standardization and the Mantel-Haenszel method. Applied Measurement in Education, 2, 217-233.

Hambleton, R.K., Swaminathan, N., and Rogers, H.J. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage.

Holland, P. W. (1985). On the study of differential item difficulty without IRT. Proceedings of the Military Testing Association, October.

Holland, P. W. & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. (Research Report No. 86-31). Princeton: Educational Testing Service.

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), Test validity. Hillsdale, NJ: Lawrence Erlbaum Associates.

Linn, R. L. and Harnisch, D. (1984). Interactions between item content and group membership on achievement test items. Journal of Educational Measurement, 18, 108-118.

Linn, R. L., Levin, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. Applied Psychological Measurement, 5, 159-173.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

McPeek W. M., & Wild C. L. (1986). Performance of the Mantel-Haenszel statistic in a variety of situations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Phillips, A. & Holland, P. W. (1987). Estimators of the variance of the Mantel-Haenszel log-odds-ratio estimate. Biometrics, 43, 425-431.

Reckase, M. D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. Journal of Educational Statistics, 4, 207-230.

Ramsey, J. O. (1989, October). Discussant comments to Theoretical and simulation-based results for Mantel-Haenszel: Choice of matching variable, power, and comparison of methods. Paper presented at the Differential Item Functioning: Theory and Practice Conference at Educational Testing Services, Princeton, NJ.

Scheuneman, J.D. & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. Journal of Educational Measurement, 27, 109-131.

Schmitt, A. (1988). Language and cultural characteristics that explain differential item funl^T^U^\ ng. Journal of Educational Measurement 25, 1-13.

Schmitt, A. and Dorans, N. (1990). Differential item functioning for minority examinees on the SAT. Journal of Educational Measurement, 27, 67-81.

Schmitt, A., Holland, P. and Dorans, N. (1989, October). Evaluating hypotheses about differential item functioning. Paper presented at the Differential Item Functioning: Theory and Practice Conference at Educational Testing Services, Princeton, NJ.

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. Journal of Educational Measurement, 22, 77-105.

Swarminathan, H. and Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27, 361-370.

Wright, D.J. An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. In A. P. Schmitt & N. J. Dorans (eds.), Differential item functioning on the Scholastic Aptitude Test (RM-87-1). Princeton, NJ:Educational Testing Service.Wright,

Zwick, R. & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. Journal of Educational Measurement, 26, 55-66.

1. "A" items have MH D-DIF values that have an absolute value of less than 1 and are not significantly different than 0 (p = .05). Any item with MH D-DIF values that are significantly different than 0 (p=.05) and have either an absolute value of 1 but less than 1.5 or absolute value of 1 but not significantly greater than 1 (p=.05) are classified as "B" items. "C" items have MH-DIF values of at least 1.5 and are significantly greater than 1 (p=.05). (See Zwick & Ericikan, 1989).

Table 1

Mantel Haenszel Analyses Design

| Analysis | Samples Code | Groups Compared | N |
|---|---|---|---|
| 1 | WW | White-focal | 354 |
| | | White | 3542 |
| 2 | WMW | White-matched focal | 168 |
| | | White | 3728 |
| 3 | AAW | AA-focal | 338 |
| | | White | 3896 |
| 4 | EOPAW | AA,EOP-focal | 166 |
| | | White | 3896 |
| 5 | NAW | AA,not EOP-focal | 172 |
| | | White | 3896 |
| 6 | EOPW | EOP-focal | 279 |
| | | White | 3915 |

Note. AA is African-American, EOP is Educational Opportunity Program.

Table 2

Advanced Math Exam: Descriptive Statistics

| Group | Mean | Minimum | Maximum | Standard Deviation |
|---|---|---|---|---|
| White (Total group) | 22.34 | 4 | 40 | 7.34 |
| White (unmatched) | 22.87 | 6 | 40 | 7.23 |
| White (matched) | 13.24 | 4 | 27 | 4.33 |
| African-American (Total group) | 15.61 | 4 | 34 | 5.81 |
| African-American (EOP) | 13.24 | 4 | 27 | 4.34 |
| African-American (Non-EOP) | 17.93 | 5 | 34 | 6.16 |
| EOP | 13.23 | 0 | 33 | 4.88 |

Note. EOP is Educational Opportunity Program.

Table 3

Descriptive Statistics for MH D-Dif:
Advance Mathematics Placement Examination

| Sample | Mean | S.D. |
|---|---|---|
| WW | .017 | .36 |
| WMW | -.006 | .53 |
| AAW | .026 | .56 |
| EOPAW | .020 | .75 |
| NAW | .020 | .53 |
| EOPW | .030 | .66 |

Table 4

Descriptive Statistics for MH D-Dif by Item Category

| Item Type | WW | WMW | AAW | EOPAW | NAW | EOPW | n |
|---|---|---|---|---|---|---|---|
| Pure | | | | | | | |
| Mean | .005 | -.071 | .095 | .089 | .075 | .064 | 24 |
| S.D. | .37 | .61 | .49 | .64 | .47 | .53 | |
| Verbal | | | | | | | |
| Mean | -.052 | .085 | -.737 | -1.21 | -.241 | -1.13 | 3 |
| S.D. | .08 | .19 | .50 | .71 | .65 | .46 | |
| Neutral | | | | | | | 13 |
| Mean | .055 | .091 | .075 | .174 | -.021 | .236 | |
| S.D. | .39 | .41 | .60 | .75 | .63 | .68 | |

Note. n= number of items.

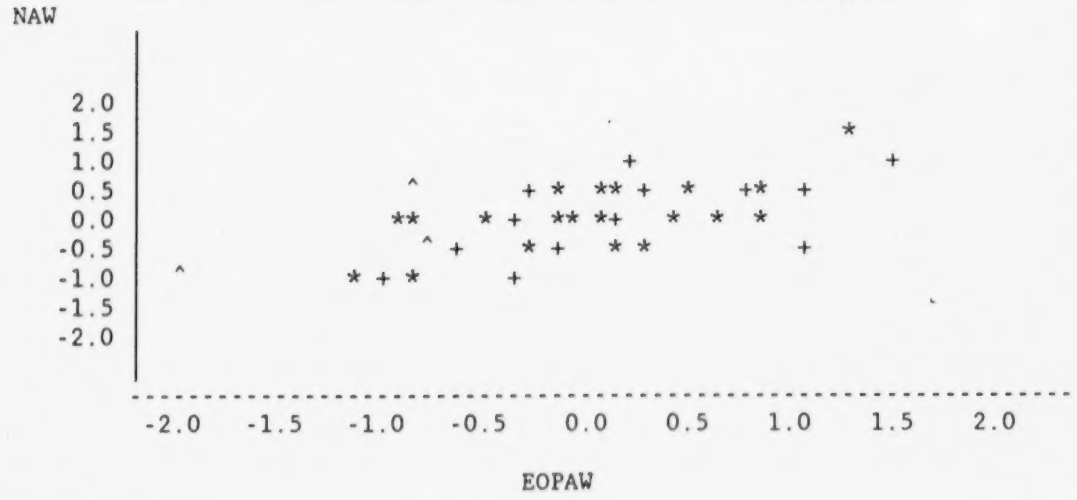
Table 5

MH D-DIF Values and MH Category Across All Analyses

| | WW | WMW | AAW | NAW | EOPAW | EOPW |
|---|---|---|---|---|---|---|
| 1 | 0.504 | 0.472 | -0.179 | -0.261 | -0.160 | 0.142 |
| 2 | -0.335 | -0.075 | 0.023 | -0.006 | 0.058 | 0.297 |
| 3 | 0.558 | 0.182 | 0.060 | -0.072 | 0.087 | -0.029 |
| 4 | -0.776 | 0.266 | -0.958 | -0.963 | -1.019 B | -0.841 |
| 5 | -0.300 | -0.277 | 0.445 | 0.011 | 0.848 | 0.349 |
| 6 | -0.316 | 0.938 | 0.245 | 0.095 | 0.426 | -0.048 |
| 7 | -0.040 | -0.104 | -0.255 | 0.505 | -0.825 | -1.391 B |
| 8 | 0.048 | -0.232 | 0.160 | 0.265 | 0.089 | 0.212 |
| 9 | -0.358 | 0.119 | -0.596 | -0.839 | -0.376 | -0.541 |
| 10 | 0.021 | 0.093 | -0.710 | -0.575 | -0.772 | -0.609 |
| 11 | -0.038 | -0.463 | 0.305 | 0.669 | -0.131 | -0.222 |
| 12 | -0.038 | -0.462 | -0.334 | -0.193 | -0.493 | 0.174 |
| 13 | 0.683 | 0.604 | 0.493 | 0.493 | 0.522 | 0.403 |
| 14 | -0.620 | -1.340 B | -0.282 | 0.063 | -0.859 | -0.572 |
| 15 | -0.649 | -1.149 B | 0.419 | 0.309 | 0.518 | 0.480 |
| 16 | -0.337 | -0.982 | 0.367 | 0.568 | 0.100 | 0.429 |
| 17 | 0.122 | 0.230 | 0.543 | 0.384 | 0.760 | 0.668 |
| 18 | -0.139 | 0.267 | -1.247 B | -0.654 | -2.028 B | -1.417 B |
| 19 | 0.213 | 0.598 | -0.069 | -0.033 | -0.084 | -0.390 |
| 20 | 0.519 | -0.596 | 0.009 | -0.255 | 0.284 | -0.146 |
| 21 | 0.227 | 0.893 | -0.288 | -0.194 | -0.356 | 0.061 |
| 22 | -0.170 | -0.025 | 0.297 | -0.015 | 0.624 | 0.325 |
| 23 | 0.497 | 0.313 | 0.646 | 0.266 | 1.086 B | 0.965 |
| 24 | 0.148 | 0.499 | -0.289 | 0.018 | -0.945 | -0.390 |
| 25 | 0.443 | -0.102 | -0.617 | -0.633 | -0.650 | -0.312 |
| 26 | 0.180 | 0.200 | 0.288 | 0.314 | 0.306 | 0.103 |
| 27 | 0.002 | 0.452 | -0.818 | -0.759 | -0.850 | -0.429 |
| 28 | -0.215 | -0.010 | 0.034 | 0.275 | -0.291 | -0.531 |
| 29 | 0.266 | 0.600 | -0.296 | -0.360 | -0.283 | -0.472 |
| 30 | -0.221 | -0.780 | 0.109 | 0.144 | 0.169 | 0.397 |
| 31 | 0.237 | -0.043 | 0.602 | 0.381 | 0.847 | 0.448 |
| 32 | 0.656 | 0.629 | 1.323 B | 1.303 B | 1.305 B | 1.217 B |
| 33 | 0.140 | -0.608 | 0.613 | 0.298 | 1.083 | 1.044 B |
| 34 | 0.047 | 0.399 | 0.375 | 0.549 | 0.169 | 0.268 |
| 35 | -0.291 | -0.440 | -0.308 | -0.626 | 0.117 | -0.069 |
| 36 | -0.245 | -0.382 | 0.197 | -0.661 | 1.060 B | 1.181 B |
| 37 | 0.241 | 0.096 | 1.239 B | 1.097 B | 1.490 B | 1.288 B |
| 38 | 0.322 | -0.131 | 0.555 | 0.796 | 0.249 | 0.482 |
| 39 | -0.007 | -0.216 | -0.942 | -0.766 | -1.175 B | -1.272 B |
| 40 | -0.292 | 0.304 | -0.128 | -0.149 | -0.114 | -0.067 |

Note. MH D-DIF values not marked with a letter were classified as "A" items.

Figure 1
Plot of MH D-DIF Values for the EOPAW and NAW Samples




+-pure item
*-neutral item
^-verbal item